

Review Session 1

Wold Decomposition, Linear Filters, and ARMA Processes

Sangmyung Ha

Fall 2026

Throughout these notes, stationarity means covariance stationarity, all processes are indexed by $t \in \mathbb{Z}$, and all random variables are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Processes are centered unless stated otherwise. Infinite random-variable series are interpreted as limits in L^2 . White noise means a mean-zero process with a constant covariance matrix and zero cross-covariances at distinct dates; independence is not required.

1 Multivariate Wold Decomposition

1.1 Scalar Hilbert spaces and linear prediction

We work in the scalar Hilbert space

$$\mathcal{H} = L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}), \quad \langle U, V \rangle = \mathbb{E}(UV), \quad \|U\|_2 = \{\mathbb{E}(U^2)\}^{1/2}.$$

Elements of $\mathcal{H}$ are measurable, square-integrable real random variables, with random variables that are equal almost surely identified. For a vector $Y = (Y_1, \dots, Y_n)'$, we use the convenient notation $\|Y\|_2^2 = \sum_{i=1}^n \|Y_i\|_2^2 = \mathbb{E}(Y'Y)$, but every projection below is a scalar projection applied separately to each coordinate.

Let $X_t = (X_{1,t}, \dots, X_{n,t})'$ be a mean-zero stationary process with $\mathbb{E}\|X_t\|^2 < \infty$. Define

$$\mathcal{H}_t^X = \overline{\text{span}}\{X_{i,s} : 1 \leq i \leq n, s \leq t\}, \tag{1.1}$$

$$\mathcal{H}^X = \overline{\text{span}}\{X_{i,s} : 1 \leq i \leq n, s \in \mathbb{Z}\}, \quad \mathcal{H}_{-\infty}^X = \bigcap_{t \in \mathbb{Z}} \mathcal{H}_t^X. \tag{1.2}$$

Here $\overline{\text{span}}$ denotes the L^2 closure of the set of finite real linear combinations. Membership in $\mathcal{H}_t^X$ therefore means that a scalar random variable can be approximated arbitrarily well in L^2 by finite linear combinations of the coordinates of current and past observations.

Let P_t be the orthogonal projection onto $\mathcal{H}_t^X$, and write

$$P_t X_s = (P_t X_{1,s}, \dots, P_t X_{n,s})'.$$

The one-step linear innovation is

$$\varepsilon_t = X_t - P_{t-1} X_t. \tag{1.3}$$

For every coordinate i , the projection theorem gives

$$\varepsilon_{i,t} \in \mathcal{H}_t^X, \quad \mathbb{E}(\varepsilon_{i,t} Y) = 0 \quad \text{for every } Y \in \mathcal{H}_{t-1}^X.$$

The predictor $P_{t-1} X_t$ is the best linear predictor; it need not coincide with the conditional expectation given the past.

1.2 Multivariate Wold Decomposition

Theorem 1.1 (Multivariate Wold decomposition with full-rank innovations). *Let $\{X_t\}_{t \in \mathbb{Z}}$ be a mean-zero, stationary, $\mathbb{R}^n$ -valued process with finite second moments. Define its one-step innovations by (1.3), and assume that*

$$\Sigma_\varepsilon := \mathbb{E}(\varepsilon_0 \varepsilon_0') \succ 0. \quad (1.4)$$

Then $\{\varepsilon_t\}$ is vector white noise with covariance Σ_ε . There are unique matrices $\{\Theta_k\}_{k \geq 0}$ such that

$$X_t = \sum_{k=0}^{\infty} \Theta_k \varepsilon_{t-k} + D_t, \quad (1.5)$$

where

$$\Theta_0 = I_n, \quad \Theta_k = \mathbb{E}(X_t \varepsilon_{t-k}') \Sigma_\varepsilon^{-1}, \quad \sum_{k=0}^{\infty} \text{tr}(\Theta_k \Sigma_\varepsilon \Theta_k') < \infty. \quad (1.6)$$

The matrices Θ_k do not depend on t , and the series in (1.5) converges in L^2 . Each coordinate of D_t belongs to $\mathcal{H}_{-\infty}^X$, and

$$\mathbb{E}(D_t \varepsilon_s') = 0 \quad \text{for all } s, t \in \mathbb{Z}.$$

The two components in (1.5) are jointly stationary and orthogonal at all leads and lags. If the process is purely nondeterministic, meaning $\mathcal{H}_{-\infty}^X = \{0\}$, then $D_t = 0$ for all t .

Proof. Step 1: Stationarity of the innovations. On finite linear combinations of the coordinates of X , define a forward shift U by

$$U \left(\sum_{(i,s) \in F} c_{i,s} X_{i,s} \right) = \sum_{(i,s) \in F} c_{i,s} X_{i,s+1},$$

where F is a finite set of coordinate–date pairs. Covariance stationarity implies that this map preserves inner products. It is therefore well defined on almost-sure equivalence classes and extends by continuity to an onto isometry of $\mathcal{H}^X$. The backward shift is its inverse.

Since $U\mathcal{H}_t^X = \mathcal{H}_{t+1}^X$ and isometry preserves orthogonality, we have

$$UP_t Y = P_{t+1} UY \quad \text{for every } Y \in \mathcal{H}^X.$$

Consequently,

$$U\varepsilon_{i,t} = U(X_{i,t} - P_{t-1}X_{i,t}) = X_{i,t+1} - P_t X_{i,t+1} = \varepsilon_{i,t+1}.$$

Thus X and ε are jointly covariance stationary. In particular,

$$\mathbb{E}(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon \quad \text{for every } t,$$

and $\mathbb{E}(X_t \varepsilon_{t-k}')$ depends only on k .

For $s < t$, every coordinate $\varepsilon_{j,s}$ belongs to $\mathcal{H}_s^X \subset \mathcal{H}_{t-1}^X$. Hence

$$\mathbb{E}(\varepsilon_{i,t} \varepsilon_{j,s}) = 0 \quad \text{for all } i, j,$$

which gives $\mathbb{E}(\varepsilon_t \varepsilon'_s) = 0$. The innovations have mean zero because the prediction spaces are generated by mean-zero random variables. They are therefore vector white noise.

Step 2: The one-step orthogonal decomposition. Define the finite-dimensional innovation space

$$\mathcal{I}_t = \text{span}\{\varepsilon_{1,t}, \dots, \varepsilon_{n,t}\}.$$

We claim that

$$\mathcal{H}_t^X = \mathcal{H}_{t-1}^X \oplus \mathcal{I}_t. \quad (1.7)$$

To see this, $\mathcal{H}_t^X$ is generated by $\mathcal{H}_{t-1}^X$ and the coordinates of X_t . For each i ,

$$X_{i,t} = P_{t-1}X_{i,t} + \varepsilon_{i,t},$$

with $P_{t-1}X_{i,t} \in \mathcal{H}_{t-1}^X$ and $\varepsilon_{i,t} \in \mathcal{I}_t$. Conversely, $\varepsilon_{i,t} = X_{i,t} - P_{t-1}X_{i,t} \in \mathcal{H}_t^X$. Moreover, $\mathcal{I}_t \perp \mathcal{H}_{t-1}^X$, and their sum is closed because it is the orthogonal sum of two closed subspaces. This proves (1.7).

Let

$$\mathcal{E}_t = \overline{\text{span}}\{\varepsilon_{i,s} : 1 \leq i \leq n, s \leq t\} \subset \mathcal{H}_t^X.$$

If $Y \in \mathcal{H}_t^X$ is orthogonal to $\mathcal{E}_t$, then (1.7) first implies $Y \in \mathcal{H}_{t-1}^X$, because $Y \perp \mathcal{I}_t$. Since also $Y \perp \mathcal{I}_{t-1}$, it follows that $Y \in \mathcal{H}_{t-2}^X$. Repeating gives $Y \in \mathcal{H}_{-\infty}^X$.

Conversely, if $Y \in \mathcal{H}_{-\infty}^X$, then $Y \in \mathcal{H}_{s-1}^X$ for every s , so Y is orthogonal to every innovation coordinate. Therefore

$$\mathcal{H}_t^X = \mathcal{H}_{-\infty}^X \oplus \mathcal{E}_t. \quad (1.8)$$

Step 3: Projection coefficients and convergence. The coordinatewise projection of X_t onto $\mathcal{I}_{t-k}$ can be written as $\Theta_k \varepsilon_{t-k}$. Its scalar orthogonality conditions, one for each pair of coordinates, are equivalent to

$$\mathbb{E}\{(X_t - \Theta_k \varepsilon_{t-k}) \varepsilon'_{t-k}\} = 0.$$

Since $\Sigma_\varepsilon \succ 0$, these conditions uniquely determine

$$\Theta_k = \mathbb{E}(X_t \varepsilon'_{t-k}) \Sigma_\varepsilon^{-1}.$$

Step 1 shows that these coefficients do not depend on t . Also,

$$\mathbb{E}(X_t \varepsilon'_t) = \mathbb{E}\{(P_{t-1}X_t + \varepsilon_t) \varepsilon'_t\} = \Sigma_\varepsilon,$$

so $\Theta_0 = I_n$.

The spaces $\mathcal{I}_t, \mathcal{I}_{t-1}, \dots$ are mutually orthogonal. Therefore

$$S_t^{(K)} = \sum_{k=0}^K \Theta_k \varepsilon_{t-k}$$

is the coordinatewise projection of X_t onto $\mathcal{I}_t \oplus \dots \oplus \mathcal{I}_{t-K}$. Pythagoras gives

$$\sum_{k=0}^K \text{tr}(\Theta_k \Sigma_\varepsilon \Theta'_k) = \|S_t^{(K)}\|_2^2 \leq \|X_t\|_2^2. \quad (1.9)$$

The nonnegative series is consequently summable. For $M > K$,

$$\|S_t^{(M)} - S_t^{(K)}\|_2^2 = \sum_{k=K+1}^M \text{tr}(\Theta_k \Sigma_\varepsilon \Theta_k') \longrightarrow 0$$

as $K, M \rightarrow \infty$. Completeness of L^2 gives a limit $S_t = \sum_{k \geq 0} \Theta_k \varepsilon_{t-k}$.

Each coordinate of S_t belongs to $\mathcal{E}_t$. For every fixed k , the residual $X_t - S_t^{(K)}$ is coordinatewise orthogonal to $\mathcal{I}_{t-k}$ whenever $K \geq k$. Taking limits shows that $X_t - S_t$ is coordinatewise orthogonal to $\mathcal{E}_t$. Thus S_t is the coordinatewise projection of X_t onto $\mathcal{E}_t$.

Step 4: The remote-past component and uniqueness. Set $D_t = X_t - S_t$. By (1.8), each coordinate of D_t belongs to $\mathcal{H}_{-\infty}^X$. In fact, $D_t = P_{-\infty} X_t$, where $P_{-\infty}$ denotes coordinatewise projection onto $\mathcal{H}_{-\infty}^X$.

For any s, t , the coordinates of D_t belong to $\mathcal{H}_{s-1}^X$, so $\mathbb{E}(D_t \varepsilon_s') = 0$. Taking L^2 limits in the series defining S_s also gives $\mathbb{E}(D_t S_s') = 0$ for every s, t .

The shift U maps $\mathcal{H}_{-\infty}^X$ onto itself and hence commutes with $P_{-\infty}$ on $\mathcal{H}^X$. Thus $U D_{i,t} = D_{i,t+1}$ and $U S_{i,t} = S_{i,t+1}$, which proves joint covariance stationarity of the two components. The remote-past component is unique by uniqueness of orthogonal projection. Finally, if $\mathcal{H}_{-\infty}^X = \{0\}$, then $D_t = 0$ for every t . $\square$

Remark (What the theorem does and does not imply). Positive definiteness gives

$$\lambda_{\min}(\Sigma_\varepsilon) \|\Theta_k\|_F^2 \leq \text{tr}(\Theta_k \Sigma_\varepsilon \Theta_k') \leq \lambda_{\max}(\Sigma_\varepsilon) \|\Theta_k\|_F^2.$$

Thus the Wold coefficients are square summable. The theorem does not assert absolute summability. Also, (1.5) gives the useful prediction formula

$$P_{t-1} X_t = D_t + \sum_{k=1}^{\infty} \Theta_k \varepsilon_{t-k}.$$

The innovation is exactly the part of X_t that cannot be predicted linearly from its past.

Remark (The full-rank assumption is a genuine restriction). Stationarity alone does not imply (1.4). For the process in Question 1 of Problem Set 1, write $X_t = (u_t, u_t + 0.5u_{t-1})'$ with $\{u_t\} \sim \text{WN}(0, \sigma^2)$ and $\sigma^2 > 0$. Its one-step innovation is $(u_t, u_t)'$, with covariance

$$\sigma^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

which is singular. The Wold decomposition can be extended to a process with reduced-rank innovations. In this case, Θ_k is not uniquely determined.

Example 1.2 (A deterministic component can be random). Let $X_t = Z + u_t$, where Z is a mean-zero, square-integrable random variable orthogonal to $\{u_t\} \sim \text{WN}(0, \sigma^2)$ with $\sigma^2 > 0$. For any fixed r ,

$$\frac{1}{N} \sum_{j=0}^{N-1} X_{r-j} = Z + \frac{1}{N} \sum_{j=0}^{N-1} u_{r-j} \longrightarrow Z \quad \text{in } L^2.$$

Hence $Z \in \mathcal{H}_r^X$ for every r , so $Z \in \mathcal{H}_{-\infty}^X$. The innovations are u_t and the Wold decomposition is $X_t = u_t + Z$. Here “deterministic” means linearly predictable from the arbitrarily remote past, not necessarily nonrandom.

2 Causality and Invertibility of MA and AR Processes

In this section, ε_t denotes a specified driving white noise. It is not automatically the one-step innovation of X_t defined as (1.3). Unless explicitly stated otherwise, its covariance is positive definite in the vector case, and its variance is $\sigma^2 > 0$ in the scalar case.

2.1 Linear processes and terminology

Consider the two-sided VMA representation

$$X_t = \sum_{j=-\infty}^{\infty} \Theta_j \varepsilon_{t-j}, \quad \sum_{j=-\infty}^{\infty} \|\Theta_j\|_F^2 < \infty, \quad (2.1)$$

where $X_t, \varepsilon_t \in \mathbb{R}^n$ and $\Theta_j \in \mathbb{R}^{n \times n}$. Orthogonality across dates implies convergence in L^2 . In particular,

$$\mathbb{E}(X_t X_t') = \sum_{j=-\infty}^{\infty} \Theta_j \Sigma_\varepsilon \Theta_j'.$$

More generally, with $\Gamma_X(h) = \mathbb{E}(X_t X_{t-h}')$,

$$\Gamma_X(h) = \sum_{j=-\infty}^{\infty} \Theta_{j+h} \Sigma_\varepsilon \Theta_j', \quad \Gamma_X(-h) = \Gamma_X(h)'. \quad (2.2)$$

The covariance series is absolutely convergent entrywise by the Cauchy–Schwarz inequality and square summability. Since it depends only on h , the process is covariance stationary.

Causality relative to the driving noise. The representation (2.1) is causal when $\Theta_j = 0$ for all $j < 0$, so that

$$X_t = \sum_{j=0}^{\infty} \Theta_j \varepsilon_{t-j}.$$

Under $\Sigma_\varepsilon \succ 0$, this is equivalent to

$$\mathbb{E}(\varepsilon_{t+k} X_t') = 0 \quad \text{for all } k > 0,$$

because

$$\mathbb{E}(\varepsilon_{t+k} X_t') = \Sigma_\varepsilon \Theta_{-k}'.$$

Thus a causal representation uses current and past, but not future, driving noise.

Invertibility of an MA representation. An MA representation is invertible if there are matrices $\{\Pi_j\}_{j \geq 0}$ such that

$$\varepsilon_t = \sum_{j=0}^{\infty} \Pi_j X_{t-j}, \quad \sum_{j=0}^{\infty} \|\Pi_j\|_F < \infty. \quad (2.3)$$

The driving noise can then be recovered through a one-sided, absolutely summable filter of the observations. This implies $\varepsilon_{i,t} \in \mathcal{H}_t^X$ for every coordinate i . The converse need not hold: membership in a closed linear span is weaker than the existence of an absolutely summable inverse filter.

Invertibility of an AR process For a VAR(p) equation

$$X_t = \sum_{j=1}^p A_j X_{t-j} + \varepsilon_t,$$

the lecture notes use *invertibility of the AR process* to mean that the AR filter can be inverted to obtain

$$X_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} \|\Psi_j\|_F < \infty. \quad (2.4)$$

This is a causal, absolutely summable MA representation. Notice the different directions: AR-invertibility expresses X_t in terms of the driving noise, whereas MA-invertibility expresses the driving noise in terms of X_t .

2.2 MA models

A finite-order vector MA model is

$$X_t = \sum_{j=0}^q \Theta_j \varepsilon_{t-j}. \quad (2.5)$$

It is causal relative to the specified noise and covariance stationary for every fixed collection of finite coefficient matrices. Stationarity does not require MA-invertibility. From (2.2),

$$\Gamma_X(h) = \sum_{j=0}^{q-h} \Theta_{j+h} \Sigma_{\varepsilon} \Theta_j' \quad (0 \leq h \leq q), \quad \Gamma_X(h) = 0 \quad (h > q).$$

Negative lags are obtained by transposition.

For a scalar MA(1),

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1},$$

the inverse filter is

$$\varepsilon_t = \sum_{j=0}^{\infty} (-\theta)^j X_{t-j}$$

when $|\theta| < 1$. The coefficients are absolutely summable, and the root of $1 + \theta z$ lies strictly outside the unit circle when $\theta \neq 0$. The case $\theta = 0$ is trivially invertible. The general MA root condition is developed in Section 3.3.

Remark (Driving noise versus innovations). For a causal, MA-invertible representation with $\Theta_0 = I_n$, the driving noise coincides with the one-step innovation. Causality makes ε_t orthogonal to the past observations, while invertibility ensures that each coordinate of $\sum_{j \geq 1} \Theta_j \varepsilon_{t-j}$ belongs to $\mathcal{H}_{t-1}^X$. Thus

$$P_{t-1} X_t = \sum_{j=1}^{\infty} \Theta_j \varepsilon_{t-j}, \quad X_t - P_{t-1} X_t = \varepsilon_t.$$

A qualification about timing. For any integer k , the process $\tilde{X}_t = \sum_{j=0}^q \Theta_j \varepsilon_{t-k-j}$ has the same autocovariance function as (2.5). Thus, autocovariances alone do not identify the time labels attached to latent driving shocks. Choosing $k = 0$ fixes this arbitrary shift, but whether a shock has a contemporaneous effect depends on its loading in Θ_0 . This timing issue should not be confused with MA-invertibility or with the definition of the one-step innovation.

2.3 AR models: the AR(1) example

An AR equation does not by itself imply stationarity. We ask whether it admits a stationary solution and, if it does, whether that solution is causal relative to the specified driving noise. Consider the scalar model

$$X_t = aX_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} \sim \text{WN}(0, \sigma^2), \quad \sigma^2 > 0. \quad (2.6)$$

Case 1: $|a| < 1$. The series

$$X_t = \sum_{j=0}^{\infty} a^j \varepsilon_{t-j} \quad (2.7)$$

converges in L^2 , since for $M > N$,

$$\mathbb{E} \left| \sum_{j=N}^{M-1} a^j \varepsilon_{t-j} \right|^2 = \sigma^2 \sum_{j=N}^{M-1} |a|^{2j} \longrightarrow 0.$$

Its limit is stationary and satisfies (2.6), as can be checked by subtracting aX_{t-1} from (2.7).

To establish uniqueness among stationary solutions, iterate the recursion for any such solution:

$$X_t = \sum_{j=0}^{N-1} a^j \varepsilon_{t-j} + a^N X_{t-N}.$$

Stationarity gives

$$\|a^N X_{t-N}\|_2 = |a|^N \|X_0\|_2 \longrightarrow 0.$$

Hence every stationary solution equals (2.7). It is AR-invertible because $\sum_{j \geq 0} |a|^j < \infty$. Notice that the remainder argument uses the uniform second-moment bound supplied by stationarity.

Case 2: $|a| > 1$. Rewrite the recursion forward as

$$X_t = a^{-1} X_{t+1} - a^{-1} \varepsilon_{t+1}.$$

Repeated substitution gives

$$X_t = - \sum_{j=1}^N a^{-j} \varepsilon_{t+j} + a^{-N} X_{t+N}.$$

For any stationary solution, $\|a^{-N} X_{t+N}\|_2 = |a|^{-N} \|X_0\|_2 \rightarrow 0$. Therefore the only possible stationary solution is

$$X_t = - \sum_{j=1}^{\infty} a^{-j} \varepsilon_{t+j}. \quad (2.8)$$

This series indeed converges in L^2 , since $\sum_{j \geq 1} |a|^{-2j} < \infty$, and its limit is stationary and satisfies (2.6).

The solution is not causal relative to this driving noise. In particular,

$$\mathbb{E}(\varepsilon_{t+1} X_t) = -\frac{\sigma^2}{a} \neq 0.$$

Consequently the AR equation is not invertible in the one-sided sense used in the course, even though it has a stationary, two-sided-filter solution. Furthermore,

$$\mathbb{E}(\varepsilon_t X_{t-1}) = -\frac{\sigma^2}{a} \neq 0,$$

so the driving noise is not the one-step innovation of X . There is no conflict with Wold's theorem, which uses innovations defined by projection, not an arbitrary driving white noise.

Case 3: $|a| = 1$. There is no finite-variance stationary solution under $\sigma^2 > 0$. Indeed, since $a \in \mathbb{R}$, this case means $a = 1$ or $a = -1$, and iteration would imply

$$\mathbb{E}|X_t - a^N X_{t-N}|^2 = \mathbb{E} \left| \sum_{j=0}^{N-1} a^j \varepsilon_{t-j} \right|^2 = N\sigma^2.$$

But stationarity and the triangle inequality imply

$$\mathbb{E}|X_t - a^N X_{t-N}|^2 \leq 4\mathbb{E}|X_0|^2,$$

which contradicts $N\sigma^2 \rightarrow \infty$.

3 How to Check Invertibility for General AR and MA Processes

3.1 Taylor and Laurent expansions

We use two facts from complex analysis.

1. If f is analytic in the disk $|z - z_0| < R$, then its Taylor expansion is

$$f(z) = \sum_{j=0}^{\infty} \frac{f^{(j)}(z_0)}{j!} (z - z_0)^j \quad (|z - z_0| < R).$$

2. If f is analytic in the annulus $r < |z - z_0| < R$, then it has a Laurent expansion

$$f(z) = \sum_{j=-\infty}^{\infty} b_j (z - z_0)^j \quad (r < |z - z_0| < R).$$

The Laurent coefficients are unique *on the specified annulus*. A function can have different Laurent expansions on different annuli. These series converge absolutely and uniformly on compact subannuli.

There is an important distinction between the following two domains:

$$\{z : |z| = 1\} \quad (\text{unit circle}), \quad \{z : |z| \leq 1\} \quad (\text{closed unit disk}).$$

If f is analytic on an open neighborhood of the closed unit disk, its Taylor expansion about zero is valid on a disk of radius greater than one. If f is analytic only on an open neighborhood of the unit circle, there is an annulus $r < |z| < R$ with $r < 1 < R$ on which it has a Laurent expansion; negative powers may be needed.

For example, for $a \neq 0$,

$$\begin{aligned} \frac{1}{1-az} &= \sum_{j=0}^{\infty} a^j z^j && \text{when } |az| < 1, \\ \frac{1}{1-az} &= -\sum_{j=1}^{\infty} a^{-j} z^{-j} && \text{when } |az| > 1. \end{aligned}$$

These two expansions correspond to the causal and noncausal AR(1) solutions. If $|a| = 1$, the pole lies on the unit circle, so neither domain contains a neighborhood of the entire unit circle.

Lemma 3.1 (Absolute summability of Laurent coefficients). *Suppose f is analytic in an annulus $r < |z| < R$ with $r < 1 < R$, and write its Laurent expansion as $f(z) = \sum_{j \in \mathbb{Z}} b_j z^j$. Then*

$$\sum_{j=-\infty}^{\infty} |b_j| < \infty.$$

Proof. Choose r_0, R_0 such that $r < r_0 < 1 < R_0 < R$. Convergence of the nonnegative-power part at $z = R_0$ implies that $b_j R_0^j \rightarrow 0$ as $j \rightarrow \infty$. Hence there is a finite constant C_+ such that

$$|b_j| \leq C_+ R_0^{-j} \quad (j \geq 0).$$

Similarly, convergence of the negative-power part at $z = r_0$ implies that $b_{-j} r_0^{-j} \rightarrow 0$, so there is a finite constant C_- such that

$$|b_{-j}| \leq C_- r_0^j \quad (j \geq 1).$$

Both bounds are geometrically summable, giving

$$\sum_{j \in \mathbb{Z}} |b_j| \leq \frac{C_+}{1 - R_0^{-1}} + \frac{C_- r_0}{1 - r_0} < \infty.$$

□

If f is analytic on a neighborhood of the closed unit disk, the same argument applies to its Taylor coefficients, and there are no negative-power terms. Thus analyticity on a neighborhood of the *circle* produces an absolutely summable two-sided filter, whereas analyticity on a neighborhood of the *closed disk* produces an absolutely summable one-sided filter.

3.2 AR(p): scalar polynomials and lag operators

Consider a scalar AR(p) equation

$$X_t = \sum_{j=1}^p \alpha_j X_{t-j} + \varepsilon_t.$$

For a stationary process, the lag operator L is first defined on finite linear combinations by $LX_t = X_{t-1}$ and then extended continuously to $\mathcal{H}^X$. Covariance stationarity makes L an onto isometry. In particular,

$$\|L^j Y\|_2 = \|Y\|_2 \quad \text{for every } Y \in \mathcal{H}^X \text{ and every } j \in \mathbb{Z}.$$

Negative powers are leads: $L^{-j}X_t = X_{t+j}$ for $j > 0$.

The AR equation becomes

$$\alpha(L)X_t = \varepsilon_t, \quad \alpha(L) = I - \sum_{j=1}^p \alpha_j L^j.$$

We associate with this operator the scalar polynomial

$$\alpha(z) = 1 - \sum_{j=1}^p \alpha_j z^j.$$

The polynomial $\alpha(z)$ is a function of a complex number; $\alpha(L)$ is an operator on random variables. A convergent coefficient expansion connects the two.

Why an absolutely summable filter is well defined. Let $\sum_{j \in \mathbb{Z}} |b_j| < \infty$, with real coefficients, and define

$$B_N(L) = \sum_{|j| \leq N} b_j L^j.$$

For $M > N$,

$$\|B_M(L) - B_N(L)\|_{\text{op}} \leq \sum_{N < |j| \leq M} |b_j| \longrightarrow 0,$$

where $\|\cdot\|_{\text{op}}$ is the operator norm induced by $\|\cdot\|_2$. Thus $B_N(L)$ converges in operator norm to a bounded operator $B(L)$ on $\mathcal{H}^X$. For every $Y \in \mathcal{H}^X$,

$$B(L)Y = \sum_{j \in \mathbb{Z}} b_j L^j Y, \quad \|B(L)Y\|_2 \leq \left(\sum_{j \in \mathbb{Z}} |b_j| \right) \|Y\|_2.$$

In particular, the filtered series belongs to the same closed space and satisfies

$$\mathbb{E} \left| \sum_{j \in \mathbb{Z}} b_j X_{t-j} \right|^2 \leq \mathbb{E} |X_0|^2 \left(\sum_{j \in \mathbb{Z}} |b_j| \right)^2. \quad (3.1)$$

A two-sided inverse when there are no unit-circle roots. Suppose

$$\alpha(z) \neq 0 \quad \text{for every } |z| = 1. \quad (3.2)$$

Because α is a finite polynomial, $1/\alpha(z)$ is analytic on an annulus containing the unit circle. Write

$$\frac{1}{\alpha(z)} = \sum_{j \in \mathbb{Z}} b_j z^j.$$

Lemma 3.1 gives $\sum_j |b_j| < \infty$. Since α has real coefficients, uniqueness of the Laurent expansion implies that the b_j are real as well.

Note that $b_j - \sum_{\ell=1}^p \alpha_\ell b_{j-\ell}$ is the coefficient on the j -th power of z in the identity $\alpha(z) \sum_j b_j z^j = 1$. The uniqueness of Laurent coefficients thus gives

$$b_j - \sum_{\ell=1}^p \alpha_\ell b_{j-\ell} = \begin{cases} 1, & j = 0, \\ 0, & j \neq 0 \end{cases}$$

Absolute convergence permits multiplication and regrouping of the corresponding operator series. Hence

$$\alpha(L)B(L) = B(L)\alpha(L) = I, \quad B(L) = \sum_{j \in \mathbb{Z}} b_j L^j.$$

This proves that $B(L)$ is actually the inverse, not merely a well-defined filter. Every stationary solution must therefore satisfy

$$X_t = \alpha(L)^{-1} \varepsilon_t = \sum_{j \in \mathbb{Z}} b_j \varepsilon_{t-j}. \quad (3.3)$$

Conversely, starting with white noise, this series constructs a stationary solution: it converges in L^2 , and the coefficient identities verify the AR equation. Thus (3.2) guarantees a unique stationary solution, possibly involving future driving noise.

A one-sided inverse when all roots are outside the closed disk. Suppose the stronger condition

$$\boxed{\alpha(z) \neq 0 \quad \text{for every } |z| \leq 1} \quad (3.4)$$

holds. Then $1/\alpha(z)$ is analytic on a neighborhood of the closed unit disk and has a Taylor expansion

$$\frac{1}{\alpha(z)} = \sum_{j=0}^{\infty} \psi_j z^j, \quad \sum_{j=0}^{\infty} |\psi_j| < \infty.$$

Consequently,

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

so the AR process is invertible.

The condition is also necessary for such an absolutely summable one-sided representation when $\sigma^2 > 0$. Indeed, substituting the representation into the AR equation and using orthogonality of the white noise forces the convolution coefficients to equal those of the identity filter. Thus, with $\psi(z) = \sum_{j \geq 0} \psi_j z^j$,

$$\alpha(z)\psi(z) = 1 \quad \text{for every } |z| \leq 1.$$

Absolute summability makes $\psi(z)$ continuous and finite on the closed unit disk, so α cannot vanish there.

Examples. For AR(1), $\alpha(z) = 1 - az$ has its finite root at $z = 1/a$ when $a \neq 0$. The one-sided condition is $|a| < 1$; $a = 0$ is trivially included. For AR(2),

$$\alpha(z) = 1 - \alpha_1 z - \alpha_2 z^2,$$

and every finite root must have modulus strictly greater than one. Roots on the unit circle are excluded. Analyticity merely on $|z| < 1$, without a neighborhood of the closed disk, is not enough.

3.3 MA(q) and ARMA root conditions

For a scalar MA(q), write

$$X_t = \beta(L)\varepsilon_t, \quad \beta(z) = 1 + \sum_{j=1}^q \beta_j z^j.$$

If

$$\boxed{\beta(z) \neq 0 \quad \text{for every } |z| \leq 1,} \quad (3.5)$$

then

$$\frac{1}{\beta(z)} = \sum_{j=0}^{\infty} \pi_j z^j, \quad \sum_{j=0}^{\infty} |\pi_j| < \infty,$$

and the operator argument gives

$$\varepsilon_t = \beta(L)^{-1} X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}.$$

Conversely, an absolutely summable one-sided inverse implies $\beta(z)\pi(z) = 1$ on the closed unit disk by the same white-noise coefficient argument as for AR(p). Hence (3.5) is necessary and sufficient for MA-invertibility.

For a scalar ARMA(p, q),

$$\alpha(L)X_t = \beta(L)\varepsilon_t,$$

assume $\alpha(0) = \beta(0) = 1$ and that α and β have no common roots. The two distinct inversions are

$$\begin{aligned} X_t &= \frac{\beta(L)}{\alpha(L)} \varepsilon_t && \text{if } \alpha(z) \neq 0 \text{ for } |z| \leq 1, \\ \varepsilon_t &= \frac{\alpha(L)}{\beta(L)} X_t && \text{if } \beta(z) \neq 0 \text{ for } |z| \leq 1. \end{aligned}$$

Each fraction denotes the corresponding absolutely summable one-sided filter. The first gives a causal MA representation; the second recovers the driving noise from the observations of a stationary solution.

3.4 A different approach: companion matrices

For the scalar AR(2), define

$$Y_t = \begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix}, \quad A = \begin{pmatrix} \alpha_1 & \alpha_2 \\ 1 & 0 \end{pmatrix}, \quad e_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}.$$

Then

$$Y_t = AY_{t-1} + e_t,$$

and repeated substitution gives

$$Y_t = \sum_{j=0}^{N-1} A^j e_{t-j} + A^N Y_{t-N}.$$

If every eigenvalue of A has modulus strictly less than one, then $\rho(A) < 1$, where $\rho(A)$ is the spectral radius. For any r with $\rho(A) < r < 1$, there is a finite constant C_r such that

$$\|A^j\|_{\text{op}} \leq C_r r^j \quad (j \geq 0).$$

This also holds when A is not diagonalizable: the polynomial factors arising from Jordan blocks are dominated by a geometric bound with ratio r .

It follows that $\sum_{j \geq 0} \|A^j\|_{\text{op}} < \infty$ and $\sum_{j \geq 0} \|A^j\|_{\text{op}}^2 < \infty$. Since the e_t are orthogonal across dates,

$$\mathbb{E} \left\| \sum_{j=N}^{M-1} A^j e_{t-j} \right\|^2 \leq \mathbb{E} \|e_0\|^2 \sum_{j=N}^{M-1} \|A^j\|_{\text{op}}^2 \longrightarrow 0.$$

Thus $\sum_{j \geq 0} A^j e_{t-j}$ converges in L^2 and defines a stationary solution. For any stationary solution, the remainder satisfies

$$\|A^N Y_{t-N}\|_2 \leq C_r r^N \|Y_0\|_2 \longrightarrow 0.$$

Consequently,

$$Y_t = \sum_{j=0}^{\infty} A^j e_{t-j}. \quad (3.6)$$

Which roots are inside and which are outside? The companion eigenvalues solve

$$\det(\lambda I_2 - A) = \lambda^2 - \alpha_1 \lambda - \alpha_2 = 0.$$

On the other hand,

$$\det(I_2 - zA) = 1 - \alpha_1 z - \alpha_2 z^2 = \alpha(z).$$

Thus a nonzero companion eigenvalue λ corresponds to the finite AR-polynomial root $z = 1/\lambda$. Therefore

$$\boxed{\rho(A) < 1 \iff \alpha(z) \neq 0 \text{ for every } |z| \leq 1.} \quad (3.7)$$

The *companion eigenvalues* must be strictly *inside* the unit circle; the *AR-polynomial roots* must be strictly *outside*.

3.4.1 The AR(2) coefficient region

For real α_1, α_2 , the equivalent coefficient restrictions are

$$\boxed{\alpha_2 > -1, \quad \alpha_1 + \alpha_2 < 1, \quad \alpha_2 - \alpha_1 < 1.} \quad (3.8)$$

Here is a direct verification. Let λ_1, λ_2 be the roots of $q(\lambda) = \lambda^2 - \alpha_1 \lambda - \alpha_2$. If both have modulus below one, then

$$|\alpha_2| = |\lambda_1 \lambda_2| < 1, \quad q(1) = 1 - \alpha_1 - \alpha_2 > 0, \quad q(-1) = 1 + \alpha_1 - \alpha_2 > 0.$$

The positivity statements hold for real roots in $(-1, 1)$ and for a complex-conjugate pair, where the products are squared moduli. These inequalities imply (3.8).

Conversely, (3.8) implies $|\alpha_2| < 1$ and $q(1), q(-1) > 0$. If the roots are nonreal, they are conjugates and satisfy $|\lambda_1|^2 = |\lambda_2|^2 = -\alpha_2 < 1$. If a real root were greater than or equal to one,

positivity of $q(1)$ would force both roots to exceed one, contradicting $|\lambda_1 \lambda_2| < 1$. The same argument using $q(-1)$ excludes a root less than or equal to minus one. Hence both roots lie in $(-1, 1)$.

Equivalently, the admissible region is $-1 < \alpha_2 < 1 - |\alpha_1|$. It is the interior of the triangle with vertices $(-2, -1)$, $(2, -1)$, and $(0, 1)$; its boundary is excluded.

3.4.2 AR(2) autocovariances

Under (3.8), let $\gamma(h) = \mathbb{E}(X_t X_{t-h})$. Since the stationary solution is causal, ε_t is orthogonal to past observations. Multiplying the AR(2) equation by X_{t-h} and taking expectations gives

$$\begin{aligned}\gamma(1) &= \alpha_1 \gamma(0) + \alpha_2 \gamma(1), \\ \gamma(2) &= \alpha_1 \gamma(1) + \alpha_2 \gamma(0), \\ \gamma(0) &= \alpha_1 \gamma(1) + \alpha_2 \gamma(2) + \sigma^2.\end{aligned}$$

Solving yields

$$\gamma(0) = \frac{\sigma^2(1 - \alpha_2)}{(1 + \alpha_2)\{(1 - \alpha_2)^2 - \alpha_1^2\}}, \quad \gamma(1) = \frac{\alpha_1}{1 - \alpha_2} \gamma(0). \quad (3.9)$$

The remaining autocovariances are obtained recursively from

$$\gamma(h) = \alpha_1 \gamma(h-1) + \alpha_2 \gamma(h-2) \quad (h \geq 2), \quad \gamma(-h) = \gamma(h).$$

The autocorrelation function is $\gamma(h)/\gamma(0)$. This connects the root conditions to the autocovariance calculations in Problem Set 1.

3.4.3 Extension to VAR(p)

For $X_t \in \mathbb{R}^n$ satisfying $X_t = A_1 X_{t-1} + \cdots + A_p X_{t-p} + \varepsilon_t$, use the block companion matrix

$$\mathbf{A} = \begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{pmatrix}.$$

For $p = 1$, simply take $\mathbf{A} = A_1$. With

$$\mathcal{A}(z) = I_n - \sum_{j=1}^p A_j z^j,$$

block elimination gives

$$\det(I_{np} - z\mathbf{A}) = \det \mathcal{A}(z).$$

Therefore the causal, absolutely summable VAR representation is obtained under

$$\boxed{\rho(\mathbf{A}) < 1 \iff \det \mathcal{A}(z) \neq 0 \text{ for every } |z| \leq 1.} \quad (3.10)$$

The resulting MA coefficients satisfy

$$\Psi_0 = I_n, \quad \Psi_j = \sum_{\ell=1}^{\min(p,j)} A_\ell \Psi_{j-\ell} \quad (j \geq 1), \quad \sum_{j=0}^{\infty} \|\Psi_j\|_F < \infty.$$

The root condition concerns the *full matrix polynomial*, or equivalently the block companion matrix, not the eigenvalues of the individual lag matrices considered separately.

For a square vector MA polynomial $\Theta(z) = \Theta_0 + \cdots + \Theta_q z^q$, the analogous sufficient and necessary condition for an absolutely summable one-sided inverse, with full-rank driving noise, is

$$\det \Theta(z) \neq 0 \quad \text{for every } |z| \leq 1.$$

This includes nonsingularity of Θ_0 . The inverse matrix function is analytic on a neighborhood of the closed unit disk, so its Taylor coefficients supply the inverse filter.

Course references

These notes accompany *Stationary Processes in the Time Domain*, Sections 2.1–2.4, and Problem Set 1 (Fall 2026). The term “invertible AR process” follows the convention in Section 2.3 of the lecture notes.